

AD-A157 486

DETERMINATION OF THE WIENER MOLECULAR BRANCHING INDEX  
FOR THE GENERAL TREE(U) GEORGIA UNIV ATHENS  
E R CANFIELD ET AL. 24 JUN 85 TR-4 N00014-84-K-0365

1/1

UNCLASSIFIED

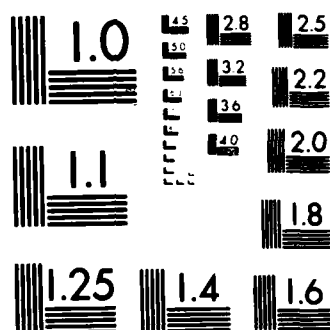
F/G 12/1

NL

END

FORMED

etc.



MICROCOPY RESOLUTION TEST CHART  
NBS 1963-A

AD-A157 486

OFFICE OF NAVAL RESEARCH

Contract N00014-84-K-0365

Task No. NR 051-861

TECHNICAL REPORT NO. 4

Determination of the Wiener Molecular Branching

Index for the General Tree

by

E.R. Canfield\*, R.W. Robinson\*, and D.H. Rouvray

Prepared for Publication

in

Journal of Computational Chemistry

University of Georgia  
Departments of Computer Science\* and Chemistry  
Athens, Georgia

June 24, 1985

Reproduction in whole or in part is permitted for  
any purpose of the United States Government

This document has been approved for public release  
and sale; its distribution is unlimited.

DTIC FILE COPY

DTIC  
ELECTE  
JUL 19 1985

S D  
G

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>Technical Report No. 4</b>	2. GOVT ACCESSION NO. <b>AD-A157486</b>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) <b>DETERMINATION OF THE WIENER MOLECULAR BRANCHING INDEX FOR THE GENERAL TREE</b>		5. TYPE OF REPORT & PERIOD COVERED <b>Technical Report</b>
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) <b>E.R. Canfield,* R.W. Robinson,* and D.H. Rouvray</b>		8. CONTRACT OR GRANT NUMBER(s) <b>N00014-84-K-0365</b>
9. PERFORMING ORGANIZATION NAME AND ADDRESS <b>University of Georgia Departments of Computer Science* and Chemistry Athens, GA 30602</b>		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>NR 051-861</b>
11. CONTROLLING OFFICE NAME AND ADDRESS <b>Office of Naval Research Department of the Navy Arlington, VA 22217</b>		12. REPORT DATE <b>June 24, 1985</b>
		13. NUMBER OF PAGES <b>39</b>
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) <b>This document has been approved for public release and sale; its distribution is unlimited.</b>		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES <b>To be published in "Journal of Computational Chemistry"</b>		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) <b>Wiener index molecular branching graph theory chemical trees</b>		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <b>The many applications of the distance matrix, D(G), and the Wiener branching index, W(G), in chemistry are briefly outlined. W(G) is defined as one half the sum of all the entries in D(G). A recursion formula is developed enabling W(G) to be evaluated for any molecule whose graph G exists in the form of a tree. This formula, which represents the first general recursion formula for trees of any kind, is valid irrespective of the valence of the vertices of G or of the degree of branching in G. Several closed expressions giving W(G) for special classes of tree molecules are derived from the general formula. One illustrative worked</b>		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

85

07

101

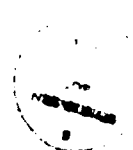
Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## 20. Abstract

example is also presented. Finally, it is shown how the presence of an arbitrary number of heteroatoms in tree-like molecules can readily be accommodated within our general formula by appropriately weighting the vertices and edges of G.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A11	



S. N 0102- LF- 014- 6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

# Determination of the Wiener Molecular Branching Index for the General Tree

E.R. Canfield<sup>a</sup>, R.W. Robinson<sup>a</sup>, and D.H. Rouvray<sup>b\*</sup>

Departments of Computer Science and Chemistry, University of Georgia,  
Athens, Georgia 30602

The many applications of the distance matrix,  $D(G)$ , and the Wiener branching index,  $W(G)$ , in chemistry are briefly outlined.  $W(G)$  is defined as one half the sum of all the entries in  $D(G)$ . A recursion formula is developed enabling  $W(G)$  to be evaluated for any molecule whose graph  $G$  exists in the form of a tree. This formula, which represents the first general recursion formula for trees of any kind, is valid irrespective of the valence of the vertices of  $G$  or of the degree of branching in  $G$ . Several closed expressions giving  $W(G)$  for special classes of tree molecules are derived from the general formula. One illustrative worked example is also presented. Finally, it is shown how the presence of an arbitrary number of heteroatoms in tree-like molecules can readily be accommodated within our general formula by appropriately weighting the vertices and edges of  $G$ .

---

<sup>a</sup> In Department of Computer Science

<sup>b</sup> In Department of Chemistry

\* To whom all correspondence should be addressed.

## I. General Introduction

Contemporary chemistry is focusing to an ever-increasing extent on the relationships between the structure of molecules and their physicochemical properties. As part of this scenario, much attention has been devoted in recent years to the search for suitable graph-theoretical means of describing the topological structure of a vast array of known chemical systems. In particular, there has been widespread usage of topological graphs and matrices for the characterization of both individual molecular species and a variety of intermolecular interactions.<sup>1-3</sup> These developments have brought in their wake a resurgence of activity in applying combinatorics and graph theory to the solution of chemical problems.<sup>4-6</sup> However, although the utilization of graphs in chemistry has a time-honored and comprehensive history,<sup>7</sup> the adoption of matrices by chemists is of much more recent vintage. Matrices which have been employed to date for the characterization of chemical systems include the adjacency

CONFIDENTIAL

- 
1. D.H. Rouvray, Roy. Inst. Chem. Revs 4, 173 (1971).
  2. I. Gutman and N. Trinajstić, Fortschr. Chem. Forsch. 42, 50 (1973).
  3. D.H. Rouvray, in Chemical Applications of Graph Theory, ed. A.T. Balaban, Academic Press, London, 1977, chap. 7.
  4. D.H. Rouvray and A.T. Balaban, in Applications of Graph Theory, ed. R.J. Wilson and L. Beineke, Academic Press, London, 1979, chap. 7.
  5. N. Trinajstić, Chemical Graph Theory I, CRC Press, Boca Raton, Florida, 1983, chap. 4.
  6. N. Trinajstić, Chemical Graph Theory II, CRC Press, Boca Raton, Florida, 1983.
  7. D.H. Rouvray, In preparation.

matrix, the incidence matrix, the cycle matrix, and the distance matrix.<sup>2-5</sup> Our prime focus of interest here will center on the distance matrix, and more especially on its derivation for the important class of graphs commonly referred to as chemical trees.<sup>8</sup> These graphs have been used extensively in the physical and biological sciences for the depiction and characterization of many different structures and systems. Following chemical intuition, the usual restriction placed upon the subclass of chemical trees known as carbon trees (C-trees) has been that no vertex in the tree could have a valence exceeding four. In our derivation of the distance matrix for trees, however, no restriction of any kind has been imposed upon the types of trees that can be considered. Accordingly, we are now able to present the first, closed recursion formula for the general tree. Our formula remains valid irrespective of the valences of the vertices, the degree of branching present, or the presence of weighted vertices or edges in the tree.

The distance matrix,  $D(G)$ , can be defined for any graph  $G$  constructed on a total of  $n$  vertices.<sup>9</sup>  $D(G)$  is the real, symmetric matrix of order  $n \times n$  whose elements,  $d_{ij}$ , are defined as follows:

$$D(G) = \begin{cases} d_{ii} = 0 \\ d_{ij} = \text{length of shortest path between} \\ \text{vertices } i \text{ and } j \text{ in } G \end{cases}$$

As we are concerned here only with tree graphs, the number of vertices,  $n$ , will always exceed the number of edges in  $G$  by one. Moreover, the

---

8. For the first use of this term see: A. Cayley, Phil. Mag. 47, 444 (1874).

9. D.M. Cvetković, M. Doob, and H. Sachs, Spectra of Graphs, Academic Press, New York, 1980.



path connecting any pair of vertices in  $G$  will always be unique. In the case of chemical trees, the diagonal  $d_{ii}$  entries are not necessarily always zero. The  $d_{ii}$  assume the value zero only if the species represented are C-trees. Similarly, in certain instances, the  $d_{ij}$  ( $i \neq j$ ) can also take nonintegral values. This circumstance arises whenever heteroatoms are present in the molecule, for then appropriately edge- and vertex-weighted trees are employed to represent the species. The corresponding  $d_{ii}$  and  $d_{ij}$  values can be obtained from general formulas put forward by Barysz et al.<sup>10</sup> In such cases, the characterization of the tree structure by the modified  $D(G)$  matrices is, of course, no longer a strictly topological one.

Up to the present, numerous applications of the distance matrix in a wide variety of different disciplines have been documented. Apart from the frequent use of  $D(G)$  as a foundation for a number of topological indices (discussed in some detail below), the matrix has also been employed in several other chemical contexts. Clark and Kettle<sup>11</sup> made use of  $D(G)$  in their study of isomer interconversions for ensembles of stereochemically nonrigid molecules such as  $PF_5$ , with the  $d_{ij}$  representing the shortest path sequence by which one isomer could be

- 
10. M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastava, and N. Trinajstić, in *Chemical Applications of Topology and Graph Theory*, ed. R.B. King, Elsevier, Amsterdam, 1983, p. 222.
  11. M.J. Clark and S.F.A. Kettle, *Inorganica Chimica Acta* 14, 201 (1975).

converted into another.  $D(G)$  has also found applications in the modelling of crystal growth<sup>12</sup> and in the study of crystal vacancies,<sup>13</sup> where the  $d_{ij}$  now represent lattice bonds. By means of this approach it was possible not only to characterize imperfections in crystals and clusters but also to predict vacancy favored positions. Another application of  $D(G)$  has involved the prediction of the mean  $\pi$ -electron energies and energy gaps in conjugated infinite polymers using both the Hückel and Pariser-Parr-Pople approximations.<sup>14</sup> More recently,  $D(G)$  has been used in the calculation of a number of physicochemical parameters for these polymers.<sup>15</sup> Biochemists have utilized  $D(G)$  in the construction of evolutionary and phylogenetic trees,<sup>16,17</sup> with the  $d_{ij}$  representing the mutation distances between genes as estimated from cytochrome c sequences. The matrix has also been employed in several other disciplines, including psychology,<sup>18</sup> communications theory,<sup>19</sup> and network flow design.<sup>20</sup> Finally, and perhaps not surprisingly,  $D(G)$  has also been the subject of intensive investigation

- 
12. D. Bonchev, O. Mekenyan, and H. Fritsche, *Cryst. Growth* 49, 90 (1980).
  13. (a) D. Bonchev, O. Mekenyan, and H. Fritsche, *Phys. Stat. Sol. A* 55, 181 (1979); (b) D. Bonchev, O. Mekenyan, and H. Fritsche, *Phys. Stat. Sol. A* 56, 607 (1979).
  14. (a) D. Bonchev and O. Mekenyan, *Z. Naturforsch.* 35A, 739 (1980); (b) D. Bonchev, O. Mekenyan, and O.E. Polansky, *Z. Naturforsch.* 36A, 643 (1981).
  15. (a) D. Bonchev, O. Mekenyan, and O.E. Polansky, *Z. Naturforsch.* 36A, 647 (1981); (b) O. Mekenyan, S. Dimitrov, and B. Bonchev, *Eur. Polym. J.* 19, 1185 (1983).
  16. W.M. Fitch and E. Margoliash, *Science* 155, 279 (1967).
  17. W.H. Li, *Proc. Natl. Acad. Sci. USA* 78, 1085 (1981).
  18. H.J. Leavitt, *J. Abn. and Soc. Psychol.* 46, 38 (1951).
  19. R.L. Graham and H.O. Pollak, *Bell Syst. Tech. J.* 50, 2495 (1971).
  20. M.L. Fredman, *SIAM J. Comput.* 5, 83 (1976).

by many mathematicians.<sup>21-26</sup>

The earliest use of the distance matrix in chemistry was in an implicit form. In 1947 Wiener<sup>27</sup> proposed the idea of using a so-called path number,  $w$ , for the purpose of correlating the molecular structure of alkane species with their boiling point, molecular refractivity, molecular volume, and heats of isomerization and vaporization. He defined  $w$  as the sum of the chemical bonds existing between all pairs of carbon atoms in such molecules. Wiener also put forward<sup>27</sup> a polarity number,  $p$ , which was defined as the number of pairs of carbon atoms separated by three carbon-carbon bonds in an alkane species. The boiling points of individual alkanes,  $t_B$ , were assumed to be given by a biparametric linear relationship of the form:

$$t_B = a \cdot w + b \cdot p + c \quad (1)$$

where  $a$ ,  $b$ , and  $c$  are constants for a given group of isomers. The

- 
21. S.L. Hakimi and S.S. Yau, Quart. Appl. Math. 22, 305 (1964).
  22. M. Edelberg, M.R. Garey, and R.L. Graham, Discrete Math. 14, 23 (1976).
  23. R.L. Graham, A.J. Hoffman, and H. Hosoya, J. Graph Theory 1, 85 (1977).
  24. R.L. Graham and L. Lovasz, Adv. in Math., 29, 60 (1978).
  25. A. Neumaier, Eur. J. Comb. 1, 163 (1980).
  26. J.M.S. Simões-Pereira and C.M. Zamfirescu, Lin. Alg. and Applic. 44, 1 (1982).
  27. (a) H. Wiener, J. Am. Chem. Soc., 69, 17 (1947); (b) H. Wiener, J. Am. Chem. Soc. 69, 2636 (1947); (c) H. Wiener, J. Chem. Phys. 15, 766 (1947); (d) H. Wiener, J. Phys. Chem. 52, 425 (1948); (e) H. Wiener, J. Phys. Chem. 52, 1082 (1948).

two structural variables,  $w$  and  $p$ , have since been shown to be directly derivable from the distance matrix,  $D(G)$ . The path number,  $w$ , is equal to half the sum of the elements of  $D(G)$ , and the polarity number,  $p$ , equals the number of off-diagonal elements in  $D(G)$  representing the distance three, i.e.

$$w(G) = \frac{1}{2} \sum_{i,j} d_{ij}(G) \quad (2)$$

$$p(G) = \frac{1}{2} \sum w_3(G) \quad (3)$$

where  $w_3(G)$  is twice the number of paths in  $G$  of length three.

Relationship (1) instituted the first use in chemistry of topological distances, i.e. path counts, for the correlation of physicochemical properties with the degree of molecular branching in individual species. Relationship (2), originally due to Hosoya,<sup>28</sup> demonstrates the close connection between  $w(G)$  and the distance matrix,  $D(G)$ . Nowadays, the Wiener path number is usually regarded as the first topological index.<sup>6</sup> Since many authors currently refer to this index by the symbol  $W(G)$ , we shall also adopt this convention hereafter. Topological indices (TIs) provide in general a convenient means of assigning a scalar magnitude to any structure of chemical significance. Starting from the chemical graph of an individual species, a TI can be derived by applying an appropriate mathematical algorithm or formula to the graph. The resulting scalar descriptor is subsequently employed for the mathematical characterization of the species in question. Over

---

28. H. Hosoya, Bull. Chem. Soc. Japan 44, 2332 (1971).

the past decade, a large number of TIs have been advocated, with special emphasis being placed on those which can offer a reliable means of reflecting the amount of branching present in both acyclic and cyclic molecular species. Attesting to the current high level of interest in this field, several reviews of TIs and their manifold applications to chemistry have recently appeared.<sup>6,29-32</sup>

The Wiener index has been widely used in a variety of guises for the correlation of branching with the bulk physicochemical properties of species. As mentioned above, Wiener himself correlated  $W(G)$  with a number of different physical properties of alkane species. This work was elaborated by Platt,<sup>33</sup> who developed an additional graphical invariant called the first neighbor sum,  $f$ , equal to the sum of the first C-C neighbors of every C-C bond in the species. The index  $f$  was used together with  $p$  and  $W(G)$  in multiparametric equations to improve on the correlations obtained by Wiener. Platt obtained very good correlations and rationalized this success by hypothesizing that  $W(G)$  provided a measure of the mean external contact area of the molecule characterized.<sup>33</sup> Altenburg<sup>34</sup> put forward a polynomial of the form:

- 
29. D. Bonchev, J.V. Knop, and N. Trinajstić, *Math. Chem.* 6, 21 (1979).
  30. I. Močoc, A.T. Balaban, O. Mekenyan, and D. Bonchev, *Math. Chem.* 13, 369 (1982).
  31. D.H. Rouvray, in *Chemical Applications of Topology and Graph Theory*, ed. R.B. King, Elsevier, Amsterdam, 1983, p. 159.
  32. A.T. Balaban, I. Močoc, D. Bonchev, and O. Mekenyan, *Topics Curr. Chem.* 114, 21 (1983).
  33. (a) J.R. Platt, *J. Chem. Phys.* 15, 419 (1947); (b) J.R. Platt, *J. Phys. Chem.* 56, 328 (1952).
  34. (a) K. Altenburg, *Kolloid Zeitschr.* 178, 112 (1961); (b) K. Altenburg, *Brennstoff Chem.* 47, 100 (1966); (c) K. Altenburg, *Brennstoff Chem.* 47, 331 (1966).

$$P(G, x) = \sum_{k=1}^n a_k x_k, \quad (4)$$

where  $a_k$  is the number of pairs of atoms in alkane species separated by the distance  $k$ , and  $x_k$  is the indexed variable. Since  $a_k$  is simply half the frequency number of the distance  $d_k$  in  $D(G)$ , it follows that:

$$W(G) = \sum_{k=1}^n a_k \cdot k \quad (5)$$

and thus another route to calculate  $W(G)$  becomes accessible.

The first extension of  $W(G)$  beyond its initial application to acyclic alkane species was made by Hosoya,<sup>28</sup> who proposed a means of calculating  $W(G)$  for cyclic alkanes. Hosoya's idea was that the  $d_{ij}$  entry in  $D(G)$  be taken as the length of the shortest path connecting the  $i$ th and  $j$ th atom in cyclic species. This idea was adopted by Rouvray<sup>35</sup> in calculating an index,  $R(G)$ , which he defined as the sum of the  $d_{ij}$  elements in  $D(G)$ . Since from equation (2) we know that  $W(G)$  is half the sum of these elements, it immediately follows that  $R(G) = 2W(G)$ . The index  $R(G)$  was determined for several different types of species, namely for members of the alkane, alkene, alkyne and arene series, and correlations were established for the boiling points, melting points, densities, refractive indices, surface tensions, and viscosities in each case. All the correlations were based on a Walker-type<sup>36</sup> relationship of the form:

---

35. (a) D.H. Rouvray, Math. Chem. 1, 125 (1975); (b) D.H. Rouvray and B.C. Crafford, South Afr. J. Sci. 72, 47 (1976).

36. J. Walker, J. Chem. Soc. 65, 725 (1894).

$$z = \alpha[R(G)]^\beta, \quad (6)$$

where  $z$  represents some physicochemical parameter, and  $\alpha$  and  $\beta$  are two constants to be determined by least squares recursion. Good to excellent correlations were found in all cases. More recently, it has been demonstrated by Bonchev et al.<sup>37</sup> that the  $W(G)$  values for monoalkyl- and o-dialkylbenzenes correlate extremely well with their chromatographic retention times. Here too, a biparametric Walker-type relationship was employed, and the results were interpreted by assuming that retention indices depend substantially upon molecular size and shape. To conclude our discussion of this topic, we recall that  $W(G)$  has also been employed to predict<sup>14</sup> the energy relationships existing in series of infinitely conjugated polymer systems (vide supra).

The index  $W(G) = \frac{1}{2} \sum d_{ij}(G)$  is important not only because of its wide variety of applications and the many excellent correlations it yields, but also because it represents the natural starting point for a number of more recent TIs. Thus TIs which are closely related to or directly derivable from  $D(G)$  include the previously mentioned polarity number of Wiener<sup>27</sup> and the first neighbor sum of Platt<sup>33</sup> as well as the Gordon and Scantlebury index,<sup>38</sup> the distance sum connectivity index of Balaban,<sup>39</sup> the information-theoretical indices of Bonchev and

- 
- 37. D. Bonchev, O. Mekenyan, G. Protić, and N. Trinajstić, *J. Chromatog.* 176, 149 (1979).
  - 38. M. Gordon and G.R. Scantlebury, *Trans. Far. Soc.* 60, 604 (1964).
  - 39. (a) D. Bonchev, A.T. Balaban, and O. Mekenyan, *J. Chem. Inf. Comp. Sci.* 20, 106 (1980); (b) A.T. Balaban, *Chem. Phys. Lett.* 89, 399 (1982).

Trinajstić,<sup>40</sup> the atomic structural index of Randić<sup>41</sup> used by Seybold to investigate the carcinogenicity of aromatic hydrocarbons<sup>42</sup> and their derivatives,<sup>43</sup> and the molecular identification number of Randić.<sup>44</sup> In view of their widespread use in chemistry and other disciplines,  $W(G)$  values have been determined for several specific classes of molecules. This has been accomplished by means of appropriate recursion formulas. The first of these, due to Hosoya,<sup>28</sup> showed that for straight chain alkane species  $W(G)$  assumes the simple closed form  $(n^3 - n)/6$ , where  $n$  is either the number of carbon atoms in the species or the number of vertices in the graph  $G$ . (We assume throughout in our discussion that all chemical species are represented by C-trees, or, equivalently, by their hydrogen-suppressed graph.) Following on from this early work,  $W(G)$  has been systematically evaluated for many systems of chemical interest in a series of papers by Bonchev and his associates. Up to the present, these workers have published recursion formulas for trees having unbranched branches,<sup>40</sup> monocyclic structures,<sup>45</sup> cyclic structures having acyclic branches,<sup>46</sup> polycyclic structures,<sup>47</sup> spiro

- 
40. D. Bonchev and N. Trinajstić, J. Chem. Phys. 67, 4517 (1977).
  41. M. Randić, Math. Chem. 7, 5 (1979).
  42. P.G. Seybold, Intn. J. Quant. Chem., Quant. Biol. Symp. 10, 95 (1983).
  43. P.G. Seybold, Intn. J. Quant. Chem., Quant. Biol. Symp. 10, 103 (1983).
  44. M. Randić, J. Chem. Inf. Comput. Sci. 24, 164 (1984).
  45. D. Bonchev, O. Mekenyan, J.V. Knop, and N. Trinajstić, Croat. Chem. Acta 52, 361 (1979).
  46. O. Mekenyan and D. Bonchev, and N. Trinajstić, Croat. Chem. Acta 56, 237 (1983).
  47. D. Bonchev, O. Mekenyan, and N. Trinajstić, Intn. J. Quant. Chem. 17, 845 (1980).



systems,<sup>48</sup> and bridged polycyclic structures.<sup>49</sup> No global recursion formula has ever been advanced for the general tree; this lacuna we now remedy here.

## II. Mathematical Formalism

The method we adopt to determine  $W(G)$  for the general tree makes inherent use of the fact that for all trees there will exist one and only one path connecting any pair of vertices in the tree. It automatically follows that this unique path must be the shortest path between the vertices. In our derivation we shall also exploit the recursive nature of trees by studying the decomposition of the general tree into a number of subtrees around a chosen vertex. We start from a completely general tree structure having a total of  $n$  vertices; this structure we shall designate by the symbol  $T$ . For convenience in visualizing our approach, we depict  $T$  as shown in Figure 1. One vertex in  $T$  has been arbitrarily selected as our focus of attention; this vertex we refer to as the root of the tree and assign it the symbol  $v_r$ . Emanating from  $v_r$  will be a total of  $m$  edges which terminate on vertices  $v_1, v_2, \dots, v_m$ . Each of the vertices  $v_1$  to  $v_m$  is regarded as the apical vertex of a subtree, as illustrated in Figure 1. These subtrees will be designated by the symbols  $T_1, T_2, \dots, T_m$  and we shall assume that each subtree contains a total of  $n_1, n_2, \dots, n_m$  vertices,

---

48. O. Mekenyan, D. Bonchev, and N. Trinajstić, *Math. Chem.* 6, 93 (1979).

49. O. Mekenyan, D. Bonchev, and N. Trinajstić, *Intn. J. Quant. Chem.* 19, 929 (1981).

respectively. Since each subtree is of arbitrary size and complexity, this type of decomposition of  $T$  can be carried out without loss of any generality.

To determine  $W(G)$  for  $T$ , account must be taken of the path distances separating all pairs of vertices in  $T$ . For convenience of visualization, we shall use the symbol  $q_{\underline{i}}$  to denote an arbitrary vertex selected from each subtree  $T_{\underline{i}}$ , where  $1 \leq \underline{i} \leq \underline{m}$ . Inspection of Figure 1 reveals that three differing types of distance need to be considered for the calculation of  $W(G)$ . These are: (I) distances originating from vertex  $v_{\underline{r}}$  and terminating on some other vertex in  $T$ ; (II) distances originating from a vertex  $q_{\underline{i}}$  in subtree  $T_{\underline{i}}$  and terminating in the same subtree  $T_{\underline{i}}$ ; and (III) distances originating from a vertex  $q_{\underline{i}}$  in subtree  $T_{\underline{i}}$  and terminating on a vertex  $q_{\underline{j}}$  in another subtree  $T_{\underline{j}}$ . By considering each of these distance types in turn and deriving an appropriate formula for each, we show how a recursive algorithm can be developed which yields  $W(G)$  for the general tree  $T$ . As is usual in developing such algorithms, we shall assume that certain information is known about each of the subtrees  $T_1$  to  $T_{\underline{m}}$ . In particular, we assume that for each subtree,  $T_{\underline{i}}$ , the sums

$$S_{\underline{i}} = \sum_{q_{\underline{i}} \in T_{\underline{i}}} d(v_{\underline{i}}, q_{\underline{i}}) \quad (7)$$

are known, where  $d(v_{\underline{i}}, q_{\underline{i}})$  is the distance traversed in moving from vertex  $v_{\underline{i}}$  to vertex  $q_{\underline{i}}$  with  $1 \leq \underline{i} \leq \underline{m}$ . We show how this information can be exploited to attain the desired result for  $T$ , as well as the corresponding quantity  $S$  for  $T$ .

### III. Derivation of $W(G)$ for the General Tree

The recursion formula we derive for the general tree,  $T$ , is comprised of three parts, one for each of the three different distance types referred to above. We now consider each distance type individually.

#### (a) Type (I) Distances

For the subtree  $T_{\underline{i}}$ , the sum of the path lengths from vertex  $v_{\underline{r}}$  to vertex  $q_{\underline{i}}$  may be expressed as:

$$s = \sum_{q_{\underline{i}} \in T_{\underline{i}}} d(v_{\underline{r}}, q_{\underline{i}}). \quad (8)$$

If all  $\underline{m}$  subtrees attached at vertices  $v_{\underline{i}}$  ( $1 \leq \underline{i} \leq \underline{m}$ ) are taken into consideration, the total number of distances becomes:

$$s(I) = \sum_{\underline{i}=1}^{\underline{m}} \left( \sum_{q_{\underline{i}} \in T_{\underline{i}}} d(v_{\underline{r}}, q_{\underline{i}}) \right). \quad (9)$$

Since  $q_{\underline{i}} \in T_{\underline{i}}$ , we know that in all cases

$$d(v_{\underline{r}}, q_{\underline{i}}) = d(v_{\underline{i}}, q_{\underline{i}}) + 1. \quad (10)$$

Accordingly, equation (9) will assume the form:

$$s(I) = \sum_{\underline{i}=1}^{\underline{m}} \left( \sum_{q_{\underline{i}} \in T_{\underline{i}}} (d(v_{\underline{i}}, q_{\underline{i}}) + 1) \right) \quad (11)$$

$$= \sum_{\underline{i}=1}^{\underline{m}} \sum_{q_{\underline{i}} \in T_{\underline{i}}} d(v_{\underline{i}}, q_{\underline{i}}) + \sum_{\underline{i}=1}^{\underline{m}} \sum_{q_{\underline{i}} \in T_{\underline{i}}} 1. \quad (12)$$

By making use of equation (7) and the definition of  $n_{\underline{i}}$ , we can rewrite equation (12) as:

$$s(I) = \sum_{\underline{i}=1}^{\underline{m}} s_{\underline{i}} + \sum_{\underline{i}=1}^{\underline{m}} n_{\underline{i}} \quad (13)$$

#### (b) Type (II) Distances

The path lengths here relate to vertices entirely located within a given subtree,  $T_{\underline{i}}$ . As pointed out in the previous section, the assumption is made that the Wiener index,  $W_{\underline{i}}$ , is known for each of the  $T_{\underline{i}}$ . Accordingly, the expression for the sum,  $s(II)$ , will take the simple form:

$$s(II) = \sum_{\underline{i}=1}^{\underline{m}} W_{\underline{i}} \quad (14)$$

#### (c) Type (III) Distances

To determine the path length sum for paths starting in one subtree,  $T_{\underline{i}}$ , and terminating in another subtree,  $T_{\underline{j}}$ , we shall focus for simplicity on the vertices  $q_{\underline{i}} \in T_{\underline{i}}$  and  $q_{\underline{j}} \in T_{\underline{j}}$ . The distance traversed in proceeding from  $q_{\underline{i}}$  to  $q_{\underline{j}}$  will be:

$$d(q_{\underline{i}}, q_{\underline{j}}) = d(q_{\underline{i}}, v_{\underline{i}}) + d(q_{\underline{j}}, v_{\underline{j}}) + 2 \quad (15)$$

Summing for all possible pairs  $(q_{\underline{i}}, q_{\underline{j}})$  by making use of the relationship

expressed in equation (15), we obtain for the general summation:

$$\begin{aligned}
 s^{(III)} = & \sum_{q_1 \in T_1} d(q_1, v_1) (n_2 + n_3 + \dots + n_{\underline{m}}) \\
 & + \sum_{q_2 \in T_2} d(q_2, v_2) (n_1 + n_3 + \dots + n_{\underline{m}}) + \dots \\
 & + \sum_{q_{\underline{m}} \in T_{\underline{m}}} d(q_{\underline{m}}, v_{\underline{m}}) (n_1 + n_2 + \dots + n_{\underline{m}-1}) \\
 & + 2 \text{ (number of pairs)}
 \end{aligned} \tag{16}$$

A closed expression for the number of pairs,  $(q_i, q_j)$ , may be derived in the following way:

$$\begin{aligned}
 \text{Number of pairs} = & n_1 n_2 + n_1 n_3 + \dots + n_1 n_{\underline{m}} \\
 & + n_2 n_3 + n_2 n_4 + \dots + n_2 n_{\underline{m}} \\
 & + \dots + n_{\underline{m}-1} n_{\underline{m}} \\
 = & [(n_1 + n_2 + \dots + n_{\underline{m}})^2 - n_1^2 - n_2^2 - \dots - n_{\underline{m}}^2] / 2 \\
 = & [(n - 1)^2 - \sum_{i=1}^{\underline{m}} n_i^2] / 2
 \end{aligned} \tag{17}$$

Equation (16) may now be simplified by making use of equation (7) thus:

$$\begin{aligned}
 s^{(III)} = & S_1 (n - 1 - n_1) + S_2 (n - 1 - n_2) + \dots + S_{\underline{m}} (n - 1 - n_{\underline{m}}) \\
 & + (n - 1)^2 - \sum_{i=1}^{\underline{m}} n_i^2 \\
 = & \sum_{i=1}^{\underline{m}} S_i (n - 1 - n_i) + (n - 1)^2 - \sum_{i=1}^{\underline{m}} n_i^2
 \end{aligned} \tag{18}$$

(d) The Final Result

At this point we are in a position to sum the topological distances represented by equations (13), (14), and (18) to obtain the result for  $W(G)$  for the general tree as follows:

$$W(G) = s(I) + s(II) + s(III) \quad (19)$$

Upon substituting into equation (19) the relevant expressions and carrying out some simplifications of the resulting equation, we obtain our final formula for  $W(G)$ , which we display here together with two other formulas which will be needed to recursively determine the Wiener index:

$$n(G) = 1 + \sum_{i=1}^m n_i \quad (20)$$

$$s(G) = n(G) - 1 + \sum_{i=1}^m s_i \quad (21)$$

$$W(G) = \sum_{i=1}^m [s_i (n - n_i) + w_i - n_i(n_i - 1)] + (n - 1)^2 \quad (22)$$

IV. Derivation of  $W(G)$  for Certain Classes of Molecules

We now demonstrate that the completely general formula presented in equation (22), can be readily applied to certain specific classes of molecules to derive closed recursion formulas.

### The Star Molecule

For a molecule possessing a graph  $G$  in the form of a star, as illustrated in Figure 2,  $W(G)$  assumes a particularly simple form. Since each of the subtrees is now represented by a single vertex, the terms in equation (22) have simple numerical values:

$$n_i = 1$$

$$s_i = 0$$

$$w_i = 0$$

The equation for the star molecule thus simplifies to:

$$W(G) = (n - 1)^2, \quad (23)$$

a result originally obtained by Bonchev and Trinajstić.<sup>40</sup>

### The Unbranched Chain Molecule

As mentioned above, the unbranched chain path molecule also has a  $W(G)$  index which can be obtained as a simple closed formula. If one end of the chain is taken as the root, as illustrated in Figure 3, the graph  $G$  can be decomposed into one subtree containing  $(n - 1)$  vertices. The terms in equation (22) are then given by the following expressions, assuming by induction the indicated expression for  $W_1$ :

$$n_1 = (n - 1)$$

$$s_1 = 1 + 2 + \dots + (n - 2) = \frac{1}{2} (n - 1) (n - 2)$$

$$w_1 = \frac{1}{6} [(n - 1)^3 - (n - 1)]$$

Substitution of these expressions into equation (22), yields the result for the unbranched chain molecule:

$$W(G) = \frac{1}{6} (n^3 - n) \quad (24)$$

in accordance with the results of Hosoya,<sup>28</sup> and Bonchev and Trinajstić.<sup>40</sup>

#### The Star Molecule with Attached Chain

In this case we take as the root the vertex at the end of the chain which is attached to the star, as shown in Figure 4. Let  $\ell$  be the length of the chain, so that the root vertex has degree  $n - \ell$ . Labeling the subtree at the root which comprises part of the chain as number one, we have

$$n_1 = \ell$$

$$S_1 = 1 + 2 + \dots + (\ell - 1) = \binom{\ell}{2}$$

$$W_1 = \frac{1}{6} (\ell^3 - \ell)$$

$$n_{\underline{i}} = 1$$

$$S_{\underline{i}} = 0 \quad \text{for } \underline{i} = 2, \dots, (n - \ell)$$

$$W_{\underline{i}} = 0$$

as in the two cases treated already. Then from (22) we calculate

$$W(G) = (n - 1)^2 + n \binom{\ell}{2} - \frac{1}{3} (2\ell + 5) \binom{\ell}{2} \quad (25)$$



$$\text{or } W(G) = \binom{\ell}{2} \left[ n - \frac{2\ell + 5}{3} \right] + (n - 1)^2$$

Note that (23) is obtained when  $\ell = 1$  and (24) when  $\ell = n - 1$ .

### The Star Molecule with Attached Chains

Here we assume that the spokes forming the star have lengths  $\ell_1, \dots, \ell_k$  as illustrated in Figure 5. By taking the central vertex as the root, the graph  $G$  can again be decomposed into  $k$  subtree chains of length  $\ell_i$ , for  $i = 1, \dots, k$ , where  $n = \sum_{i=1}^k \ell_i + 1$ . The terms in equation (22) will be:

$$\begin{aligned} n_i &= \ell_i \\ S_i &= 1 + 2 + \dots + (\ell_i - 1) = \binom{\ell_i}{2} \\ W_i &= \frac{1}{6} (\ell_i^3 - \ell_i) \end{aligned}$$

and these lead to the following expression for  $W(G)$ :

$$W(G) = \sum_{i=1}^k \binom{\ell_i}{2} \left[ n - \frac{2\ell_i + 5}{3} \right] + (n - 1)^2. \quad (26)$$

The special case of this formula when all the chains are of length two, i.e.  $\ell_i = 2$ , is given by the equation:

$$W(G) = k(n - 3) + (n - 1)^2$$

Since in this instance  $k = (n - 1)/2$ , the formula can be further simplified to:

$$W(G) = \frac{1}{2}(n - 1)(3n - 5). \quad (27)$$

# The Tree Molecule with Unbranched Branches

Here we let the branches of the tree molecule be of arbitrary length, as illustrated in Figure 6. Let us assume that the branches have lengths  $L_1, L_2, \dots, L_k$  which are attached to the main chain at positions  $1 < l_1 \leq \dots \leq l_k < C$ , where  $C$  denotes the length of the main chain. We shall establish the formula

$$\begin{aligned} W(G) = & \binom{C+1}{3} + \sum_{\underline{i}} \binom{L_{\underline{i}}+1}{3} + C \cdot \sum_{\underline{i}} \binom{L_{\underline{i}}+1}{2} \\ & + \sum_{\underline{i}} L_{\underline{i}} \left[ \binom{l_{\underline{i}}}{2} + \binom{C-l_{\underline{i}}+1}{2} \right] + \sum_{\underline{i} < \underline{j}} L_{\underline{i}} L_{\underline{j}} (l_{\underline{j}} - l_{\underline{i}}) \\ & + \sum_{\underline{i} < \underline{j}} [L_{\underline{j}} \binom{L_{\underline{i}}+1}{2} + L_{\underline{i}} \binom{L_{\underline{j}}+1}{2}] \end{aligned} \quad (28)$$

by a double induction first on  $\underline{m}$ , the number of branches, and then on  $L_1$ , the number of atoms in the first branch. To start, when  $\underline{m} = 0$  the equation reduces to the formula  $\binom{C+1}{3}$ , established earlier for a chain of length  $C$ . Now assume  $\underline{m} > 0$  and that  $L_1 \geq 1$ . We apply (22), using the last atom of branch 1 as the root. In this case there is only one subtree, so we have

$$n_1 = (n - 1).$$

When only one subtree is present, equation (22) for  $W$  becomes

$$W = S_1 + W_1 + (n - 1)$$

$$= W_1 + S.$$

By induction, we know that  $W_1$  is given by (28) with  $L_1$  replaced by

$(L_1 - 1)$ . The difference between the original (28) and (28) with  $L_1$  replaced by  $(L_1 - 1)$  is (accounting here for the six terms respectively):

$$\begin{aligned} & 0 + [(\binom{L_1 + 1}{3}) - (\binom{L_1}{3})] + C \cdot [(\binom{L_1 + 1}{2}) - (\binom{L_1}{2})] \\ & + [(\binom{\ell_1}{2}) + (\binom{C - \ell_1 + 1}{2})] + \sum_{j>1} L_j (\ell_j - \ell_1) \\ & + \sum_{j>1} [L_j ((\binom{L_1 + 1}{2}) - (\binom{L_1}{2})) + (\binom{L_j + 1}{2})], \end{aligned}$$

which, after some elementary algebra, is the same as:

$$\begin{aligned} & 0 + (\binom{L_1}{2}) + C \cdot L_1 + [(\binom{\ell_1}{2}) + (\binom{C - \ell_1 + 1}{2})] \\ & + \sum_{j>1} L_j (\ell_j - \ell_1) + \sum_{j>1} [L_j L_1 + (\binom{L_j + 1}{2})]. \end{aligned}$$

Our proof is completed by showing that this difference agrees with S. By a direct computation we have

$$\begin{aligned} S &= 1 + 2 + \dots + (L_1 - 1) && \text{(points in same branch)} \\ &+ L_1 && \text{(point of juncture)} \\ &+ (L_1 + 1) + (L_1 + 2) + \dots + (L_1 + \ell_1 - 1) && \text{(points left of juncture)} \\ &+ (L_1 + 1) + (L_1 + 2) + \dots + (L_1 + (C - \ell_1)) && \text{(points right of juncture)} \\ &+ \sum_{j>1} \sum_{b=1}^{L_j} [L_1 + (\ell_j - \ell_1) + b] && \text{(points on other branches)} \\ &= (\binom{L_1 + 1}{2}) + (\ell_1 - 1) \cdot L_1 + (\binom{\ell_1}{2}) + (C - \ell_1) \cdot L_1 \\ &+ (\binom{C - \ell_1 + 1}{2}) + \sum_{j>1} [L_j L_1 + (\ell_j - \ell_1) \cdot L_j + (\binom{L_j + 1}{2})]. \end{aligned}$$

where  $\underline{b}$  is the distance from the main chain to the various atoms on the  $\underline{j}$ th branch. Again, using elementary algebra, this latter expression is seen to agree with our previously computed difference between the original (28) and (28) with  $L_1$  replaced by  $(L_1 - 1)$ . Thus, the proof is completed. Notwithstanding the fact that this is a new result, it can be shown to be equivalent to the formula published by Bonchev and Trinajstić,<sup>40</sup> though it is simpler in format.

We now apply (28) to a few special cases which have been considered individually in the literature.

#### The Tree Molecule with Branches of Length One

The graph of the molecule is illustrated in Figure 7. Again letting the length of the chain be  $C$ , we have, in terms of the notation of (28), that  $\underline{m} = 2(C - 2)$ ,  $L_1 = L_2 = \dots = L_{\underline{m}} = 1$ ,  $\ell_1 = 2$ ,  $\ell_2 = 2$ ,  $\ell_3 = 3$ ,  $\ell_4 = 3$ , ...,  $\ell_{\underline{m}} = C - 1$ , and so

$$\begin{aligned} W(G) &= \binom{C+1}{3} + 0 + C \cdot \sum_{\underline{i}=1}^{\underline{m}} 1 + \sum_{\ell=2}^{C-1} \left[ \binom{\ell}{2} + \binom{C-\ell+1}{2} \right] \cdot 2 \\ &\quad + \sum_{2 \leq \ell_1 < \ell_2 \leq C-1} (\ell_2 - \ell_1) \cdot 4 + \sum_{1 \leq \underline{i} < \underline{j} \leq \underline{m}} 2 \\ &= \frac{1}{2} (3C^3 - 27C + 32). \end{aligned} \tag{29}$$

#### The Tree Molecule with Branches of Given Length

This case is similar to the immediately preceeding case, but now all the branches are of length  $L$  instead of length one. The graph is shown in Figure 8. Again with  $C$  denoting the length of the chain,

we can use (28) with  $\underline{m} = 2(C - 2)$ ,  $L_1 = L_2 = \dots = L_{\underline{m}} = L$ ,  $\ell_1 = 2$ ,  $\ell_2 = 2$ ,  $\ell_3 = 3$ ,  $\ell_4 = 3$ ,  $\dots$ ,  $\ell_{\underline{m}} = C - 1$ , to obtain

$$\begin{aligned}
 W(G) &= \binom{C+1}{3} + \sum_{i=1}^{\underline{m}} \binom{L+1}{3} + C \cdot \sum_{i=1}^{\underline{m}} \binom{L+1}{2} \\
 &\quad + L \cdot \sum_{\ell=2}^{C-1} \left[ \binom{\ell}{2} + \binom{C-\ell+1}{2} \right] \cdot 2 + L^2 \cdot \sum_{2 \leq \ell_1 < \ell_2 \leq C-1} (\ell_2 - \ell_1) \cdot 4 \\
 &\quad + L \binom{L+1}{2} \cdot \sum_{1 \leq i < j \leq \underline{m}} 2 \\
 &= \binom{C+1}{3} + (2C - 4) \cdot \left[ \binom{L+1}{3} + C \cdot \binom{L+1}{2} + (2C - 5) \cdot L \cdot \binom{L+1}{2} \right] \\
 &\quad + 4L \left[ \binom{C}{3} + L \binom{C-1}{3} \right]. \tag{30}
 \end{aligned}$$

#### V. An Illustrative Worked Example

The formulas derived above relate to specific classes of tree-like molecules. When the Wiener index,  $W(G)$ , is required for a molecule not belonging to one of these classes, resort must be made to our general formula given in equation (22). Below we illustrate how our formula is applied in the case of the molecule of 2,5-dimethyl-2-ethyl-4-propyl hexane. This molecule has a graph,  $G$ , which does not belong in any of the above special classes; its hydrogen-suppressed graph is illustrated in Figure 9.

As in all the previously described special cases, to determine  $W(G)$  for this molecule one sets out by arbitrarily selecting one vertex in the graph of the molecule and calling this the root. In Figure 9 the root vertex we have selected is indicated by a full circle.

Starting from this root, one moves out progressively toward the extremities of the graph, as indicated by the arrows placed on the edges of the graph. It is the outermost vertices which are tackled first. For each vertex the triplet  $(\underline{n_i}, \underline{S_i}, \underline{W_i})$  is evaluated in turn starting from the outermost vertices and working back towards the root. The value of the triplet  $(\underline{n_i}, \underline{S_i}, \underline{W_i})$  for the root, which will be the last triplet to be determined, automatically gives the required value of  $W(G)$  for the molecule as a whole. This is equivalent to calculating  $W(G)$  by substituting appropriate terms into equation (22) in a recursive manner.

Because every vertex at an extremity of the graph  $G$  will have only one attached edge, the triplet  $(\underline{n_i}, \underline{S_i}, \underline{D_i})$  will always assume the value  $(1,0,0)$  for such vertices and may be written in directly. All other triplets need to be calculated. Considering next the two longest side chains in the molecule, it is readily seen that for a vertex attached to the end vertex the triplet will assume the value  $(2,1,1)$ . The third vertex from the end along the longest chain has the triplet  $(3,3,4)$ . This is calculated by summing the two previously evaluated  $\underline{n_i}$  values, and by making use of equations (20-22) as shown below:

$$\underline{n_i} = 1 + n_1 = 1 + 2 = 3$$

$$\underline{S_i} = n - 1 + S_1 = 3 - 1 + 1 = 3$$

$$\begin{aligned} \underline{W_i} &= S_1 (n - n_1) + W_1 - n_1(n_1 - 1) + (n - 1)^2 \\ &= 1 \cdot (3 - 2) + 1 - 2 \cdot 1 + 2^2 = 4 \end{aligned}$$

$$(\underline{n_i}, \underline{S_i}, \underline{W_i}) = (3,3,4).$$

The other triplets are progressively evaluated by an analogous procedure until the root vertex is reached. The values obtained for each of the triplets are shown in Figure 9. The triplet for the root comes out to be (13,27,258), so that  $W(G) = 258$ .

# VI. The Substitution of Carbon Atoms by Heteroatoms

We now consider the changes that need to be made in our general formula (22) when heteroatoms are substituted for carbon atoms in tree-like molecules. In purely graph-theoretical terms, such substitution corresponds to a weighting of the vertices and edges in the tree graph,  $G$ . The number of edges and vertices in  $G$  which can be weighted is arbitrary; if necessary, all of the edges and vertices can be weighted.

Let us suppose initially that just one heteroatom,  $A$ , is substituted for a carbon atom in the tree-like molecule and that  $A$  has the same valence as that of the atom it substitutes. The situation which then arises is illustrated in Figure 10. The initial edge weights for all edges emanating from the original carbon atom we represent as  $\delta_1, \delta_2, \dots, \delta_m$  and the new edge weights after the substitution of the carbon atom by  $A$  we shall represent as  $\alpha_1, \alpha_2, \dots, \alpha_m$ . We shall also choose the vertex associated with atom  $A$  as the root vertex. From equations (13) and (14), it follows that:

$$S_{\text{new}} = S_{\text{old}} + \sum_i n_i (\alpha_i - \delta_i) \quad (31)$$

$$W_{\text{new}} = W_{\text{old}} + n_i (S_{\text{new}} - S_{\text{old}}) - \sum_i (\alpha_i - \delta_i) n_i^2 + d(v_A, v_A)/2$$

$$= W_{old} + \sum_i (\alpha_i - \delta_i) n_i (n - n_i) + d(v_A, v_A)/2 \quad (32)$$

where  $S_{new}$  and  $W_{new}$  refer to the respective values of  $S$  and  $W$  after the substitution of atom  $A$ ,  $S_{old}$  and  $W_{old}$  refer to these values before the substitution of atom  $A$ , and  $d(v_A, v_A)$  is the weight assigned to atom  $A$  for the  $d_{ii}$  matrix entry in  $D(G)$ .

From equation (32), it is seen that the substitution of one carbon atom by one heteroatom,  $A$ , will produce a change in the value of  $W(G)$  for the tree-like molecule equal to:

$$\Delta W(G) = \sum_i (\alpha_i - \delta_i) n_i (n - n_i) + d(v_A, v_A)/2 \quad (33)$$

The effect of substituting one heteroatom is thus to bring about a change in the value of  $(\alpha_i - \delta_i)$  for each pair of vertices,  $(X, Y)$ , such that the shortest path from  $X$  to  $Y$  includes the edge in question, plus the addition of one half of the atomic weighting term,  $d(v_A, v_A)$ . The total number of pairs of vertices  $(X, Y)$  will be given by the expression  $n_i(n - n_i)$ . It follows that the substitution of  $h$  heteroatoms for carbon atoms will result in a new  $W(G)$  value for the tree-like molecule equal to:

$$W(G)_{new} = W(G)_{old} + \sum_h \sum_i (\alpha_i - \delta_i) n_i (n - n_i) + \sum_{A=1}^h d(v_A, v_A)/2 \quad (34)$$

From the above reasoning, it is evident that the substitution of heteroatoms into a tree-like molecule for which  $W(G)$  is known does not necessitate a de novo calculation of  $W(G)$ . The known  $W(G)$ , referred to above as  $W(G)_{old}$ , is simply modified upon substitution by the two



latter terms in equation (34). It need hardly be emphasized that the modification of  $W(G)_{old}$  using equation (34) involves a much shorter calculation than a recalculation of  $W(G)$  de novo. To illustrate this point, we consider the molecule shown in Figure 9 again to suppose now that the root atom is replaced by the heteroatom N. Using the tables of Barysz et al.,<sup>10</sup> we know that the  $d_{ii}$  entry for N is 0.143 and that the C-N bond has a weighting of 0.857. In this case, therefore, the new value of  $W(G)$  is given straightforwardly by the expression:

$$\begin{aligned} W(G)_{new} &= W(G)_{old} + \sum_{h=1}^3 \sum_{i=1}^3 (0.857 - 1)(13 - n_i) + \sum_{A=1} 0.143/2 \\ &= 258 - 14.586 + \frac{0.143}{2} = 243.486 \end{aligned}$$

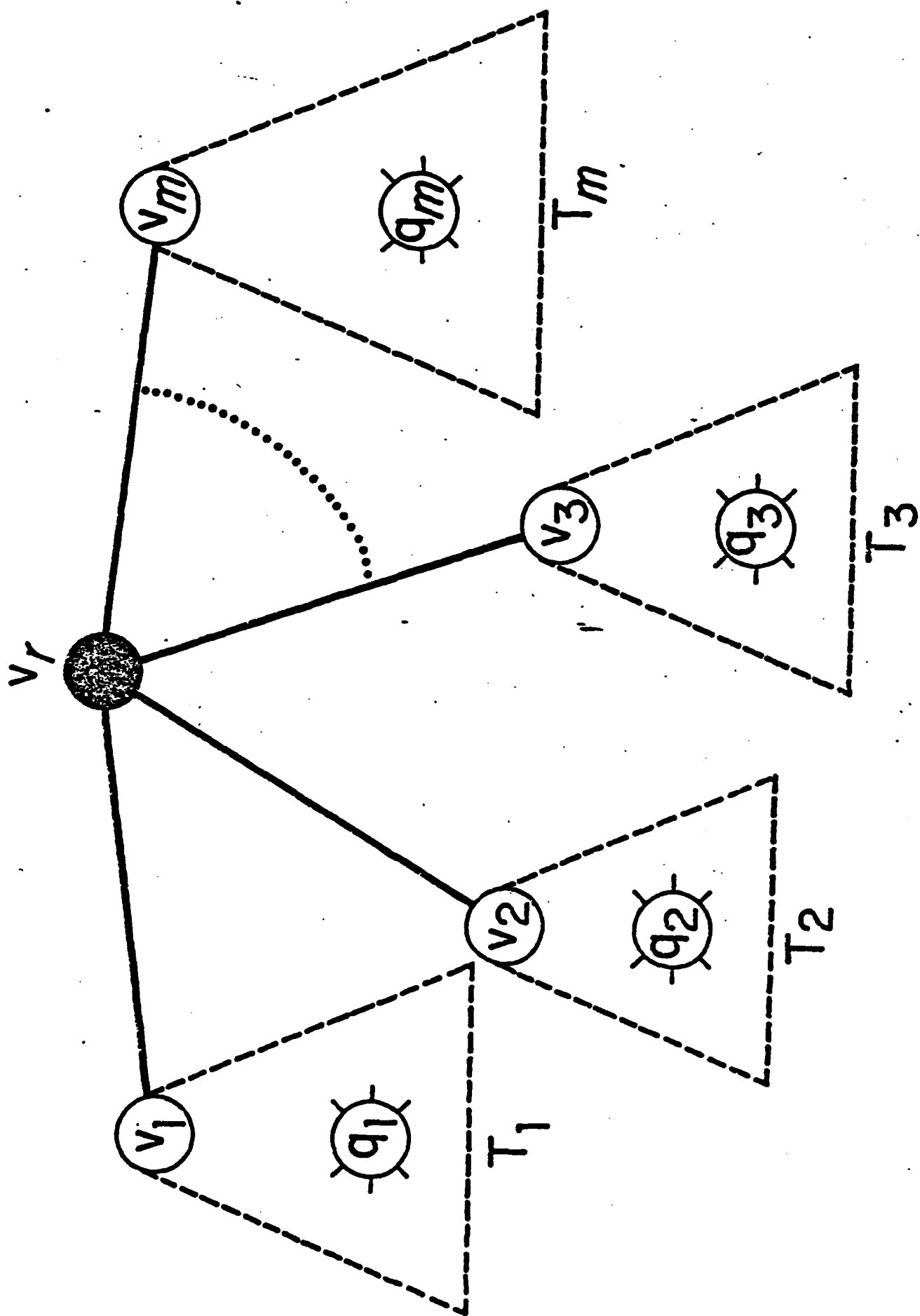
Such a procedure effects a substantial saving in time when  $W(G)_{new}$  values are to be computed.

#### Acknowledgment

We are indebted to the U.S. Office of Naval Research for partial support of this research.

### Captions to Figures

- Figure 1: A representation of the general tree, with one distinguished root.
- Figure 2: The graph of a star molecule with one distinguished vertex.
- Figure 3: The graph of an unbranched chain molecule with one distinguished vertex.
- Figure 4: The graph of a star molecule with attached chain and one distinguished vertex.
- Figure 5: The graph of a star molecule with attached chains with one distinguished vertex.
- Figure 6: The graph of a chain molecule having unbranched branches and one distinguished vertex.
- Figure 7: The graph of a tree molecule having branches of length one and one distinguished vertex.
- Figure 8: The graph of a tree molecule having branches of given length and one distinguished vertex.
- Figure 9: Graph of the molecule of 2,5-dimethyl-2-ethyl-4-propyl-hexane with one distinguished vertex.
- Figure 10: Representation of the substitution of a heteroatom A into the graph of a hydrocarbon molecule.



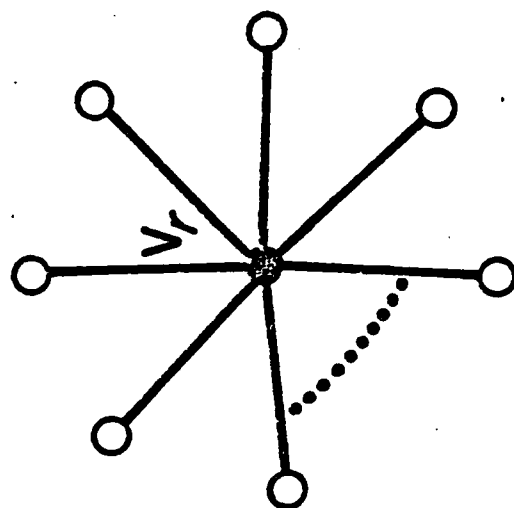




Figure 3

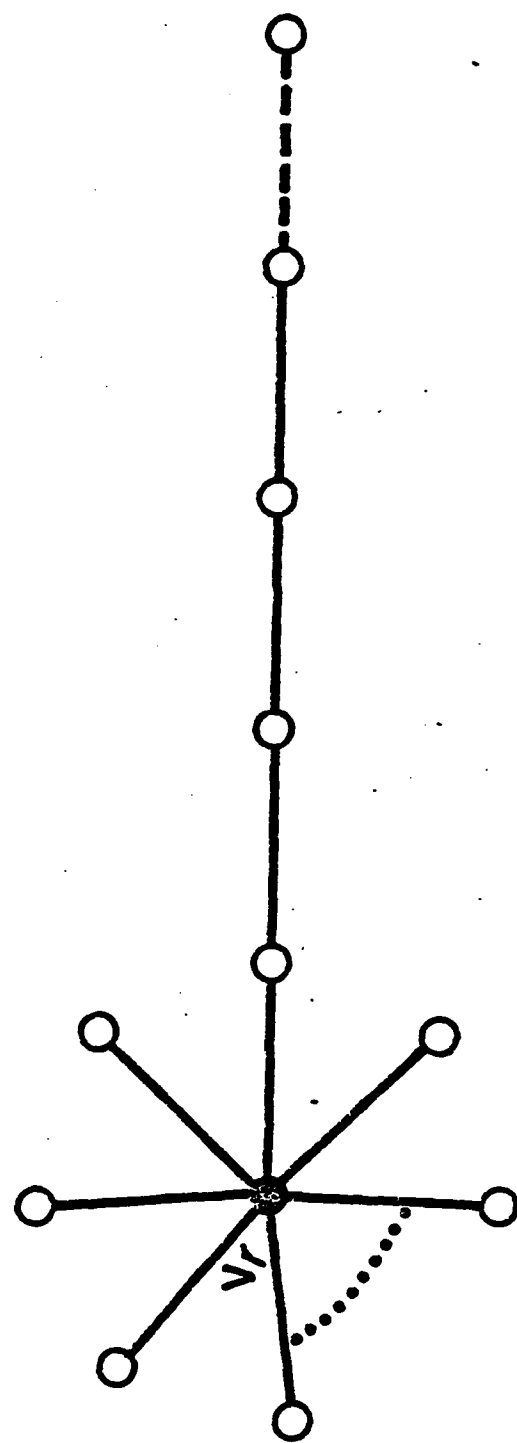


Figure 4

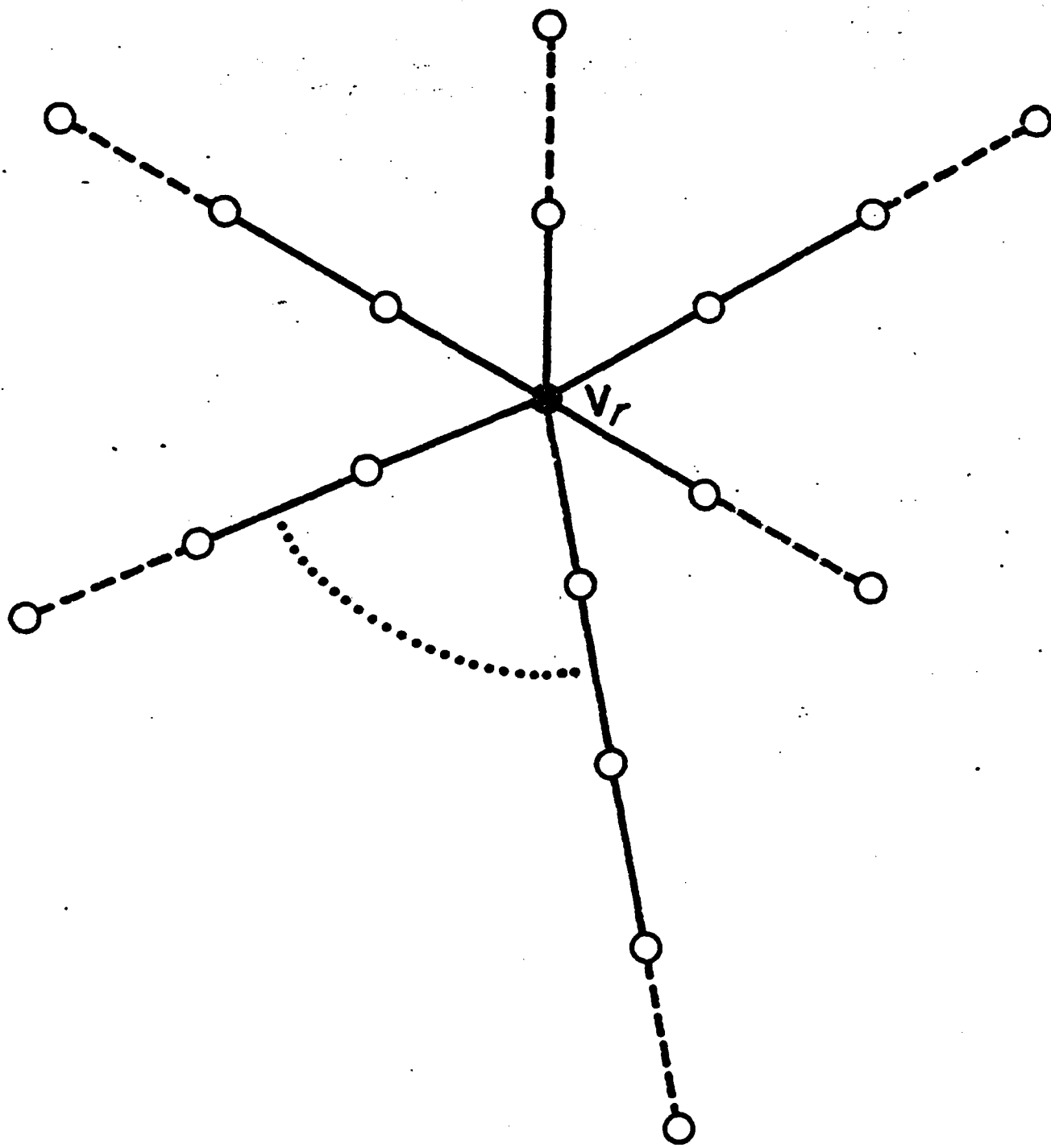
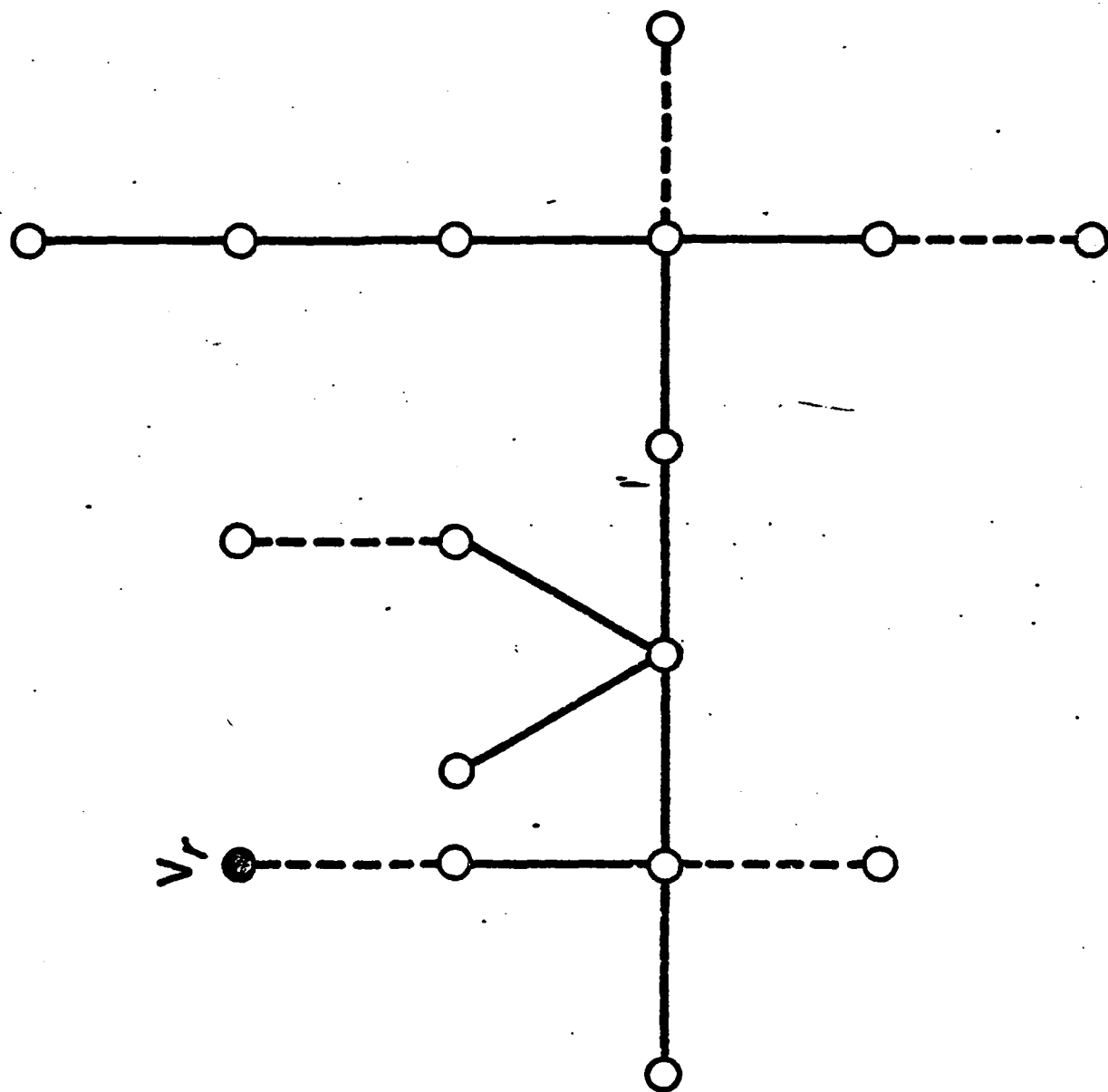


Figure 5

Figure 6





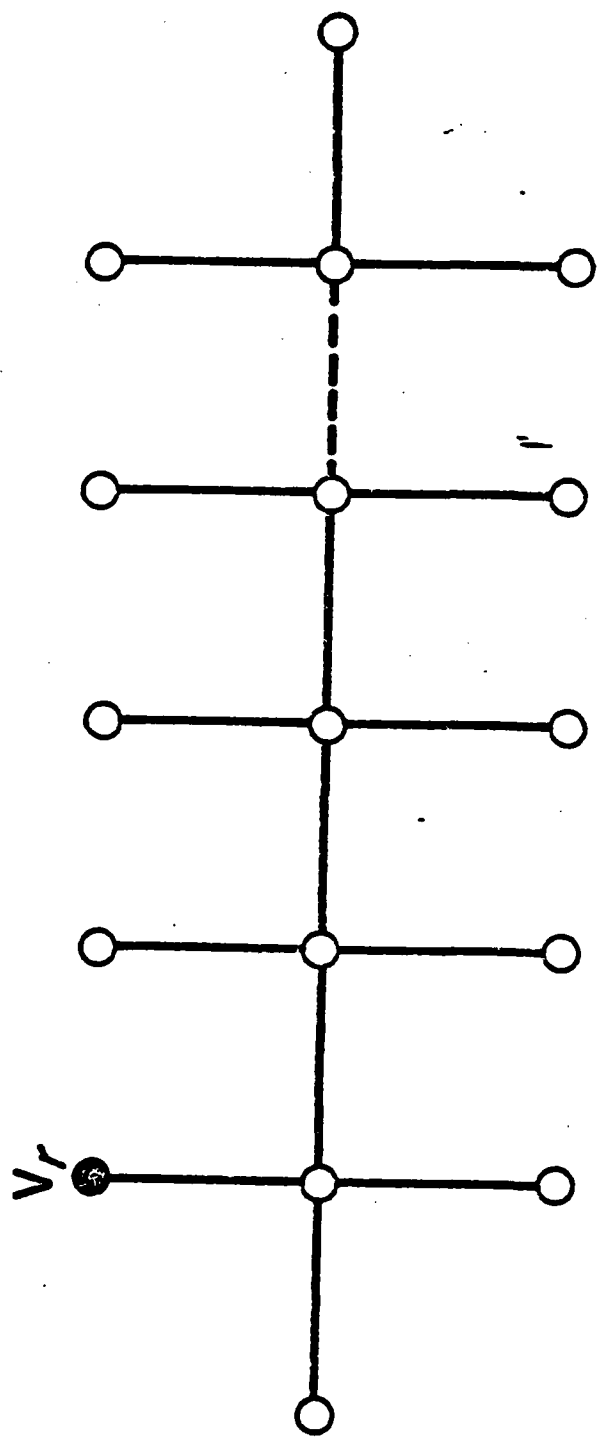


Figure 7

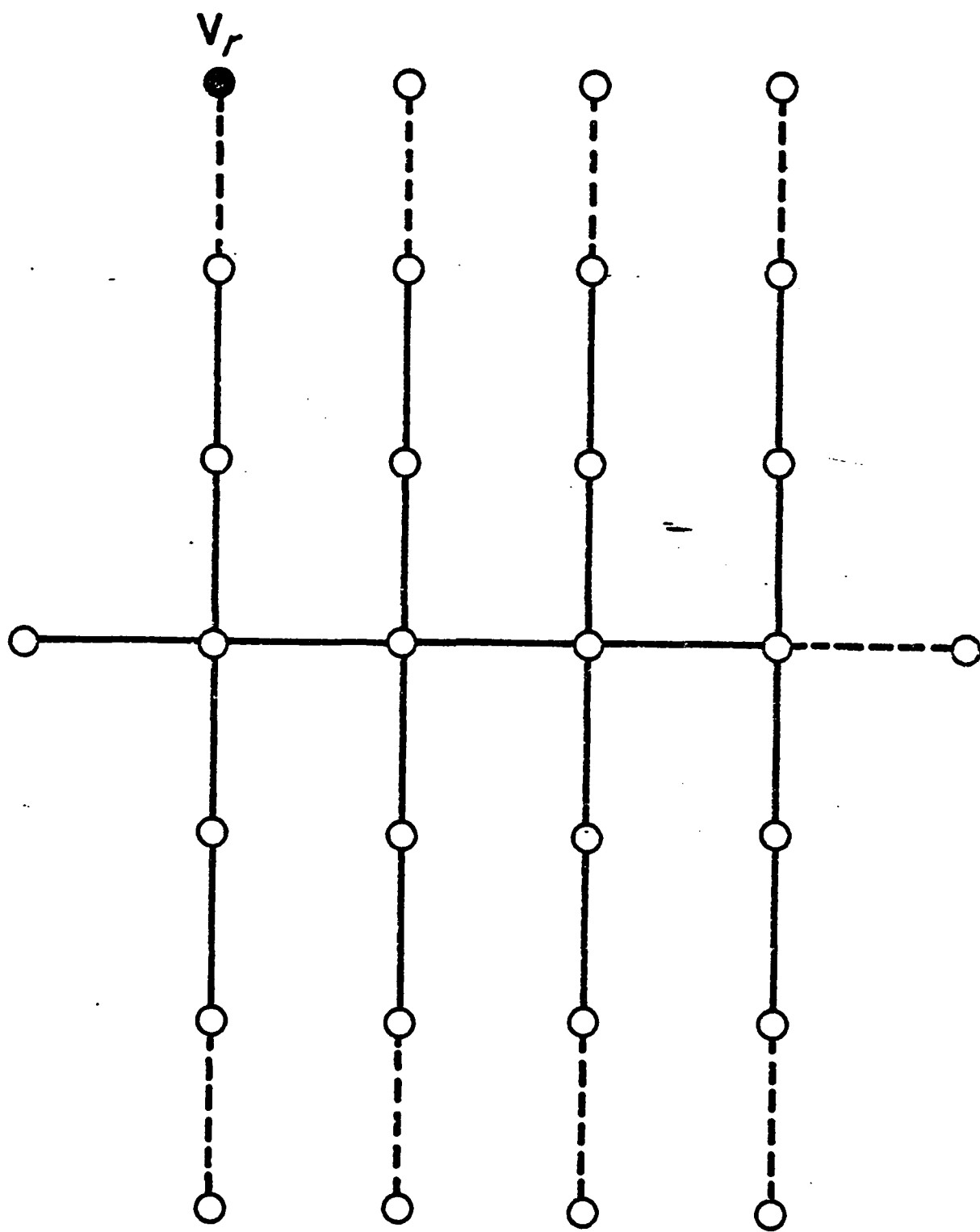


Figure 8

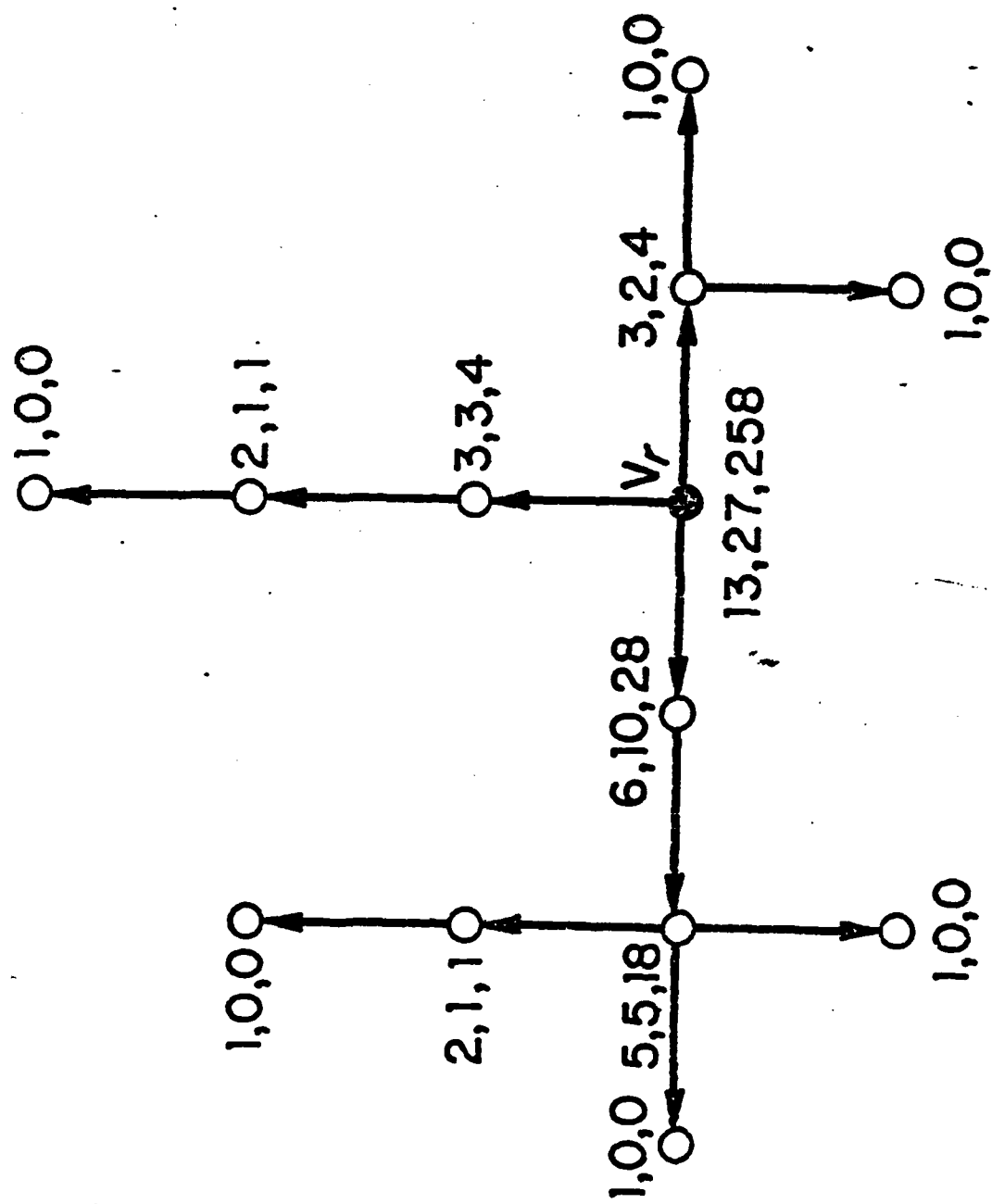


Figure 9

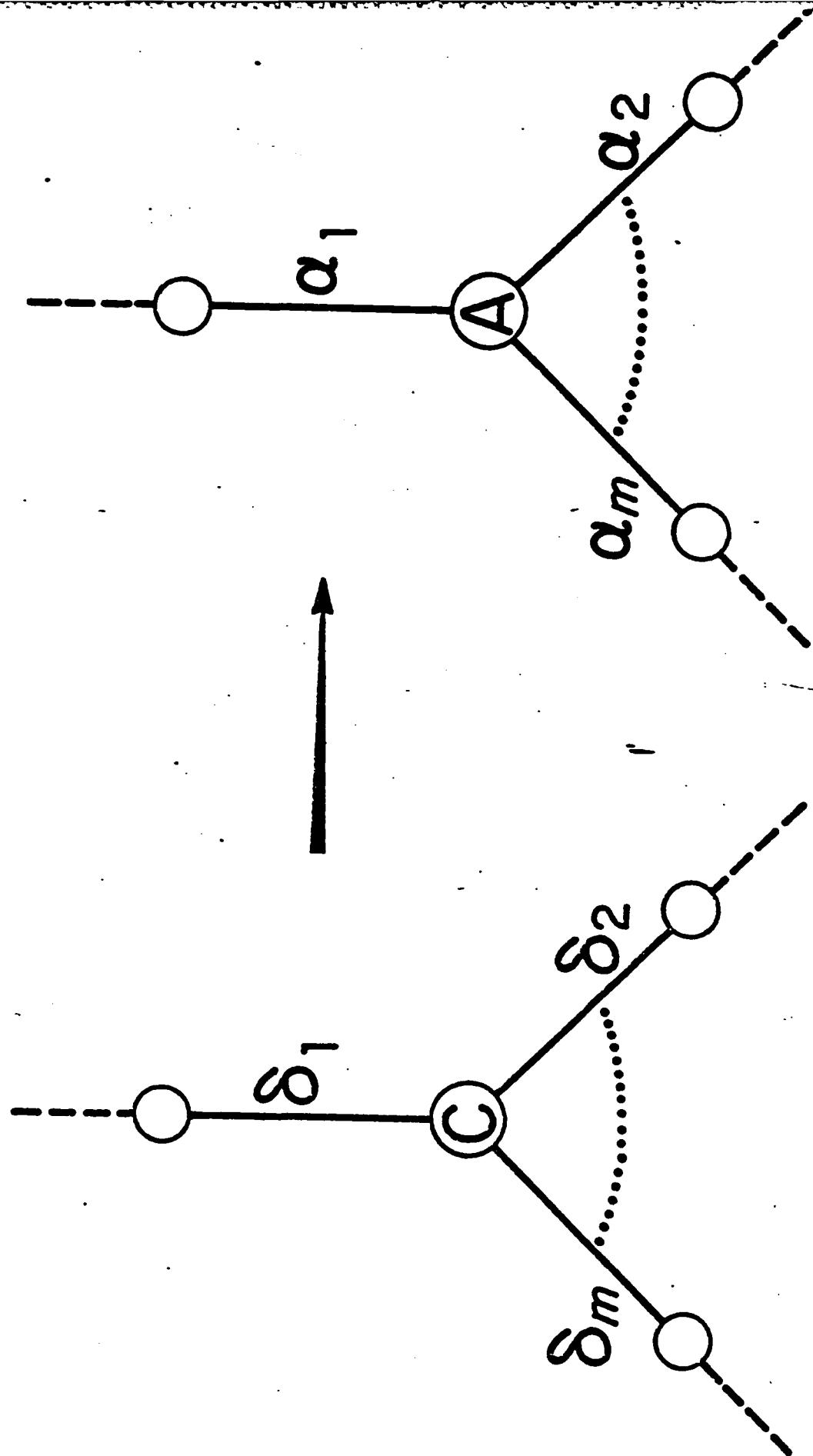


Figure 10

**END**

**FILMED**

9-85

**DTIC**